Cyber Warfare & Inadvertent Escalation

James M. Acton

The advent of cyber warfare exacerbates the risk of inadvertent nuclear escalation in a conventional conflict. In theory, cyber espionage and cyberattacks could enhance one state's ability to undermine another's nuclear deterrent. Regardless of how effective such operations might prove in practice, fear of them could generate escalatory "use-'em-before-you-lose-'em" pressures. Additionally, cyber threats could create three qualitatively new mechanisms by which a nuclear-armed state might incorrectly conclude that its nuclear deterrent was under attack. First, cyber espionage could be mistaken for a cyberattack. Second, malware could accidentally spread from systems that supported non-nuclear operations to nuclear-related systems. Third, an operation carried out by a third party could be misattributed by one state in a bilateral confrontation to its opponent. Two approaches to risk reduction are potentially viable in the short term : unilateral restraint in conducting potentially escalatory cyber operations, and bilateral or multilateral behavioral norms.

Cyber weapons may be relatively new, but non-nuclear threats to nuclear weapons and their command, control, communication, and intelligence (C3I) systems are not. In fact, before the United States dropped the bomb on Hiroshima in August 1945 – before it even conducted the world's first nuclear test in July of that year – it had started to worry about non-nuclear threats to its nascent nuclear force, in particular, Japanese air defenses.[1] As the Cold War developed, fears multiplied to encompass threats to almost every component of the United States' nuclear forces and C3I systems. While these threats emanated primarily from Moscow's nuclear forces, they were exacerbated by its improving non-nuclear capabilities, particularly in the final decade of the Cold War. A two-decade hiatus in worry following the Soviet Union's collapse is now over; today, non-nuclear threats to U.S. nuclear C3I assets – in particular, the growing capability of Chinese and Russian antisatellite weapons – are a major concern.[2]

The United States' experience is the norm. All nuclear-armed states have felt, and continue to feel, similar concerns. Indeed, the last few decades have seen the emergence of new potential vulnerabilities – this time in cyberspace – as nuclear weapons and C3I systems have come to rely increasingly on digital technology. To be sure, the networks involved in nuclear operations are almost certainly among the most secure anywhere. Yet there is broad agreement among technical experts

https://doi.org/10.1162/DAED_a_01794

that perfect network security is "impossible."[3] As a result, the possibility of cyber interference with nuclear forces and C3I systems is real.

The vulnerability of nuclear forces and C3I systems creates the risk of inadvertent escalation: that is, escalation resulting from military operations or threats that are not intended to be escalatory. So-called crisis instability, for example, could arise if a state were afraid of being disarmed more or less completely in a preemptive strike by an adversary, whether or not such fears were well founded.[4] In the most extreme case, "use-'em-or-lose-'em" pressures could lead the state to employ nuclear weapons, conceivably in its own preemptive attempt to disarm its adversary, but more likely in a limited way to try to terrify the opponent into backing down. In less extreme scenarios, a state afraid of being disarmed might take steps – issuing nuclear threats, for example, or dispersing mobile nuclear forces – that raised the likelihood of nuclear use later.

This danger is likely to be exacerbated by any cyber vulnerabilities affecting nuclear forces and C3I systems. Most directly, the existence of such vulnerabilities could intensify existing fears of being disarmed – fears that are already acute in China and Russia (as well as in Pakistan and, most likely, North Korea).[5] However, because of their unique characteristics and effects, cyber threats could create at least three qualitatively new mechanisms by which a nuclear-armed state might come to the incorrect conclusion that its nuclear deterrent was under threat. First, the purpose of cyber interference could be misinterpreted. In particular, espionage could be mistaken for an attack. Second, a cyberattack could have a more significant effect than intended. Malware implanted into information technology (IT) systems associated with non-nuclear weapons could accidentally spread into more sensitive nuclear-related systems, for instance. Third, the initiator of a cyber operation could be misidentified. An operation carried out by a third party, for example, could be misattributed by one state in a bilateral confrontation to its opponent. What makes these pathways so pernicious is that the catalyst for escalation could appear to its initiator to be a relatively benign action.

To make matters worse, such pathways could lead to inadvertent escalation even if the target of the cyber interference were not afraid of being completely disarmed. Today at least, this description fits the United States. If, in a conflict against Russia, say, the United States wrongly concluded that its strategic early-warning system was under cyberattack, it might reason that Moscow was seeking to undermine U.S. missile defenses, which use early-warning data, prior to launching a nuclear attack.[6] Given that U.S. declaratory policy explicitly highlights the option of a nuclear response to non-nuclear attacks on nuclear C3I assets, such a "misinterpreted warning" might lead Washington to use nuclear weapons.[7] But even if it did not, its response, which might include nuclear threats, could still be escalatory.

My focus here is narrowly limited to inadvertent cyber threats against, or interference with, one state's nuclear forces or C3I systems by another nuclear-armed

state (C3I systems encompass not only communication capabilities, but also the intelligence, surveillance, and reconnaissance capabilities, including early warning, that would be critical to decision-making). To be sure, cyber vulnerabilities probably create other escalation risks too, though, in my judgment, they are less serious.[8] For example, while no state would likely try to detonate another's nuclear weapons, a nihilistic terrorist group might (though it is unclear whether such a group could obtain the requisite cyber capabilities). Separately, vulnerabilities associated with conventional forces or their C3I systems could increase the likelihood of a conventional war's escalating to a higher level of violence, thus making nuclear use more credible.[9]

Cyber interference with nuclear forces and C3I systems can involve two (not mutually exclusive) types of operations: espionage and attack. Cyber espionage involves collecting data from a target IT system without otherwise damaging it. A cyberattack involves undermining the operations of the target system, typically by compromising the integrity or availability of data. Cyber tools suitable for surveilling or attacking nuclear forces or C3I systems have innumerable differences from noncyber tools, which are themselves quite varied. Six of these differences are particularly salient to the risk of inadvertent nuclear escalation.

First, cyber espionage offers the potential to obtain information about an adversary's military forces and operations that cannot plausibly be obtained in any other way. By accessing an adversary's C3I systems directly, cyber tools may be capable of exfiltrating exceptionally sensitive information, such as the locations of mobile delivery systems. This is not to suggest that cyber surveillance is infallible. As a security measure, for example, a state could choose not to track the movements of its mobile delivery systems (or it could do so only approximately). Alternatively or additionally, it could try to use a cyber intrusion in its networks to feed misinformation to the adversary. In spite of these and other limitations, however, cyber espionage almost certainly offers unique advantages. For example, no practical constellation of high-resolution surveillance satellites in low Earth orbit could provide continuous coverage of a given location on Earth's surface.[10] Cyber surveillance, by contrast, may allow for continuous monitoring of an adversary's military posture.

Second, cyber weapons offer an unparalleled capability to manipulate the data that go into decision-making. Other types of weapons, by destroying or disabling sensors or communication systems, can also deny data to decision-makers. However, their use generally alerts the target to the fact it is under attack. By contrast, if a well-designed cyber weapon is used, a loss of data may appear to be, say, the result of a malfunction, potentially allowing the attacker to conduct surprise follow-on attacks. Even more significant, cyber weapons can be used to feed false

information to decision-makers. For example, the Stuxnet virus, which was reportedly developed by the United States and Israel, was designed not only to destroy centrifuges at Iran's Natanz enrichment plant, but also to hinder plant operators from discovering the cause of these failures by producing falsely reassuring readings on monitoring equipment.[11] In a similar vein, sophisticated cyber weapons offer a unique capability to shape an adversary's perception of a battlefield by feeding misinformation into C3I systems.[12] To be sure, information operations have always been a part of warfare. However, cyber weapons represent a sea change because their effects can be tailored with great precision in real time, and because they could be used to directly influence the perceptions of high-level decision-makers.

Third, cyber operations – whether conducted for espionage or offensive purposes – can present particularly significant risks of *unanticipated* collateral effects, that is, of affecting IT systems other than the intended target.[13] Noncyber weapons can, of course, lead to collateral damage. Yet such effects are inherently constrained by geography. Moreover, the likelihood of physical collateral damage can be often quantified, at least to some extent (military planners may be able to estimate, for example, the probability of an incoming weapon missing its military target and hitting a nearby civilian facility).[14] The risks of collateral effects in cyberspace are much more difficult to estimate. Minimizing such effects relies, in part, on detailed intelligence about the target network and on connections between it and other networks. Obtaining the requisite intelligence is potentially much more difficult than identifying what surrounds a target in physical space (as is verifying that the resulting picture is complete). To complicate matters further, sophisticated malware must generally be tailored to each target and, if revealed, will become ineffective once the adversary can clean its networks and fix whatever exploit was used to gain access. As a result, the effects of cyber weapons cannot usually be understood through testing, further increasing the likelihood of unanticipated collateral damage (simulations can be used but they are only as good as the available intelligence on the target).

Fourth, in peacetime, malware used to enable a cyberattack may often be inserted into an enemy's networks – but not activated – in the hope that it will remain undetected and thus can be used in a potential future crisis or conflict. (In theory, not only can a vulnerability in an operational IT system be exploited in this way, but so too could security weaknesses in the supply chain for the system's components.) Noncyber weapons, by contrast, are generally used as and when the decision to authorize a strike on a particular target is taken.[15] One consequence of this difference is that, if a state discovers dormant malware in its networks, it can be faced with the challenge of attributing it – that is, identifying which entity is responsible for its implantation – before activation. The equivalent challenge rarely arises with the kinds of noncyber weapons typically used in interstate warfare

(though it does arise in irregular warfare or counterterrorism with unexploded ordnance).

Fifth, and relatedly, cyberattacks are generally easier to conceal than other forms of attack. As a result, decision-makers may be more inclined to authorize them. In fact, if the goal is for a cyber weapon to have either a persistent effect or an effect when triggered at some future time, the malware used in the attack must remain hidden to be effective because exposure could enable the adversary to take countermeasures.

Sixth, and finally, distinguishing between offensive operations and espionage is significantly more challenging in cyberspace than in other domains.[16] To be sure, the line dividing espionage and offensive operations in physical space is not always entirely clear. Aircraft – unmanned aerial vehicles (UAVs), in particular – are used for both surveillance and offensive operations. But the distinction is much murkier in cyberspace. One challenge is that identifying the purpose of a piece of malware – understanding whether it can be used for espionage, offensive purposes, or both – can be time-consuming. In a fast-moving conflict or crisis, this process might move slower than decision-making. Moreover, even if a state quickly and confidently established that a piece of malware could be used solely for espionage, it could not be confident that whatever vulnerability was used to introduce the malware would not also be exploited for offensive purposes – at least until it had identified and fixed the vulnerability.

States can threaten each other's nuclear forces through a combination of offensive "counterforce" operations to target nuclear-weapon delivery systems preemptively, and air and missile defense operations to intercept whatever remained. The United States openly acknowledges it would seek to limit the damage it would suffer in a nuclear war.[17] Russian doctrine is believed to embrace a similar concept.[18] India may be moving in the same direction.[19]

The question of whether, in practice, a state could actually succeed in limiting the damage it would suffer in a nuclear war to an extent that decision-makers would consider meaningful is currently a subject of considerable debate.[20] However, from the perspective of inadvertent escalation, what matters is not whether damage-limitation operations would actually prove effective, but whether a potential target believes they might. In this context, Chinese and Russian fears that the United States is seeking the capabilities – non-nuclear capabilities, in particular – to negate their nuclear deterrents could prove escalatory in a crisis or conflict by generating "crisis instability," that is, pressures to use nuclear weapons before losing the capability to do so.[21] And even though the United States is not concerned today about the possibility of being disarmed, Washington appears to be less sanguine about the future, given growing threats to its C3I assets, in particular.

Cyber capabilities could contribute to damage-limitation operations in two distinct ways. First, cyber espionage could prove useful in collecting intelligence that might increase the effectiveness of counterforce attacks and air and missile defenses, especially if complemented by effective analytic tools for synthesizing large amounts of data from multiple sources.[22] If cyber espionage helped reveal the locations of mobile weapons, for example, it could enable preemptive attacks against them. And if it helped to reveal targeting data, it could assist defenses in intercepting missiles and aircraft after launch.

Second, cyber weapons could be used, alongside other capabilities, to conduct counterforce strikes. A hypothetical cyber "kill switch" that could permanently shut down an adversary's nuclear C3I systems would certainly be attractive to any state with a damage-limitation doctrine. In practice, this kind of perfect capability seems fanciful, not least because a state could find analog or even nonelectronic ways to use its own nuclear forces given enough time (in fact, some states may even prepare such means in advance). At best, therefore, a cyberattack could be a "pause button" that delayed an adversary's ability to use its nuclear weapons. Real cyber weapons are likely to be still less effective, however. All nuclear-armed states likely operate multiple C3I systems with some degree of redundancy between them. Cyber operations would probably not prove equally effective against these different systems, potentially delaying the target from using some elements of its nuclear forces for longer periods of time than others.

Even given these limitations, however, cyberattacks could still assist with damage limitation. They could buy more time for counterforce operations to attrite an opponent's nuclear forces and reduce the coherence of any retaliatory attacks, somewhat simplifying the task of air and missile defenses. Moreover, the potential for cyberattacks to shape an adversary's perceptions could prove valuable. For example, an attacker might try to "blind" its adversary's early-warning system just before launching counterforce strikes on its nuclear forces.

Just how effective cyber-enabled damage-limitation operations might prove in an actual conflict is far from clear, not least because of the difficulty of testing cyber weapons. That said, any state that has made the enormous investments necessary to develop damage-limitation capabilities is likely to spend relatively modest additional sums on developing complementary cyber tools, and it might reach a different conclusion about their potential efficacy. Even more important, from the perspective of inadvertent escalation, its potential adversaries might do so too.

China, in particular, appears to be concerned about cyber-enabled damage limitation. Summarizing the thinking of their peers on this subject, two Chinese scholars, Tong Zhao and Li Bin, have concluded that "Chinese analysts have demonstrated an acute awareness of the potential vulnerabilities of the

country's nuclear C3I system, particularly against cyber infiltrations."[23] Russian views have been less aired. In fact, a dichotomy has emerged in what little public discussion there has been. For example, three respected experts, including a former general officer in Russia's Strategic Rocket Forces, have recently played down the threat, arguing that "because the command-and-control systems of strategic nuclear forces are isolated and highly protected, they are, in all probability, not vulnerable to cyber attacks."[24] At about the same time, however, another influential Russian scholar argued that, among the emerging non-nuclear technologies that could threaten nuclear forces, "probably the most dangerous development is cyber weapons, which could be used for non-nuclear disarming and decapitating attack by completely paralysing the entire command-and-control system."[25] News reports that Russia has created cyber defense units for its nuclear forces suggest that the Russian military may be less than sanguine about the cyber threat.[26]

Fears about cyber-enabled damage limitation may be particularly pernicious because of the potential difficulty of detecting a cyberattack. A sophisticated cyberattack on nuclear forces or C3I systems could conceivably occur without being detected. In the extreme case, a state might only find out that it had been attacked when it attempted to launch nuclear weapons and discovered that its ability to do so had been impeded in some way. If a state believed that it would be unlikely to detect an ongoing cyberattack, then it could rationally conclude that it might be under attack even in the absence of attack indicators. The simple belief that an opponent had highly sophisticated cyber capabilities could, therefore, precipitate a false positive – the incorrect assessment that an attack was underway – by itself. By contrast, if a state's nuclear forces were under assault from kinetic strikes, the target would likely be aware. To be sure, it is still not entirely impossible that a state could wrongly come to believe it was under kinetic attack. Early-warning systems, for example, have produced false warnings of incoming ballistic missile strikes.[27] But mistakes of this kind could be identified once the incoming weapons ceased to exist (though the window of time before they disappeared could be particularly dangerous).

To make matters worse, a state that was concerned about its nuclear forces and C3I systems coming under cyberattack might be inclined, especially in a crisis or conflict, to interpret ambiguous indicators in the worst possible light. For example, if one of its nuclear C3I systems malfunctioned because of, say, bad design or aging components, it might wrongly attribute the failure to a cyberattack (in fact, the temptation among operators to do so might be particularly strong if they would otherwise be held responsible for an internal failure). Regardless of precisely how it arose, however, a false positive that occurred in a crisis or conflict could generate significant escalation pressures.

Concerns about the potential for cyber operations to enhance the effectiveness of damage limitation can have effects beyond generating crisis instability at a time of heightened tensions or during a conflict. In peacetime, such concerns may induce nuclear-armed states to take steps to try to ensure that nuclear weapons could be employed when duly ordered in a crisis or conflict, even at the expense of exacerbating the danger of inadvertent or unauthorized use. Concerned states, for example, could remove permissive action links – electronic "locks" designed to prevent the unauthorized use of nuclear weapons – because of the perceived danger that they could be hacked and thus subverted to prevent authorized use.[28]

Alternatively or additionally, states could make plans to predelegate the authority to use nuclear weapons down the chain of command to guard against the possibility of the communication links serving national leaders being severed. The dangers of predelegation depend, in part, on the degree of flexibility afforded to commanders in determining whether and how to use nuclear weapons. Nevertheless, certain risks are inherent in any model. A localized communications failure might be mistaken for an attack, for example, leading to inadvertent use.[29] Predelegation also increases the risk of unauthorized use because a field commander could order the use of nuclear weapons in a scenario in which he or she was not permitted to do so. This danger becomes greater as more people are granted launch authority. In this respect, cyber threats could promote a particularly dangerous form of predelegation by inducing a state to entrust launch authority to the relatively large number of lower-level officers who are capable of issuing a launch order without electronic communications.

Surveillance operations in cyberspace, even if conducted exclusively for defensive purposes, pose unique risks of escalation. Cyber surveillance of an adversary's nuclear forces can serve purposes besides damage limitation. In any dyad involving two nuclear-armed states, each has a strong incentive to monitor the status of the other's nuclear forces at all times – and particularly during a crisis or conflict – including for the exclusively defensive purpose of spotting any preparations for nuclear use. Several intelligence collection techniques, including overhead imagery and signals intelligence, are likely used for this purpose. Given the potentially unique advantages of surveillance in cyberspace, however, states may see good reason to adopt it alongside these other approaches, especially if they judge that the likelihood of cyber espionage being detected is small.

Depending on the sophistication of the malware used and the target's defenses, the true likelihood of being detected may or may not be small, but the consequences of being caught could be significant. In fact, if the target detected ongoing cyber espionage of networks associated with its nuclear forces or C3I systems,

inadvertent escalation could result from either of two concerns that are distinct from those that might plausibly be generated by other forms of surveillance.

First, even if the target of cyber interference were convinced that the operation was being conducted exclusively for the purpose of espionage, it might worry that the data being collected could be used against it in damage-limitation operations. Intelligence collection in physical space could also enable damage limitation, but it differs from cyber surveillance in one critical respect. In a crisis or conflict, a state would generally have no way of knowing whether or not countermeasures against physical surveillance (such as camouflage or concealment) had proved effective – unless its nuclear forces were successfully attacked. By contrast, if it detected an ongoing effort to collect intelligence through its C3I networks, it would know definitively that at least some of its cyber defenses had failed. This realization might lead the state to fear that attacks on its nuclear forces were imminent.

Second, because of the difficulty of rapidly distinguishing cyber espionage from a cyberattack, espionage against nuclear forces or C3I systems would risk being misinterpreted as an attack. In theory, the use of armed UAVs for surveillance of an adversary's nuclear forces could generate a similar risk. However, a state motivated by purely defensive considerations would have strong and obvious reasons not to use armed UAVs in this way.

The risks resulting from cyber espionage being mistaken as an attack would depend on who had initiated the operation and who was the target. China or Russia might assess that U.S. cyber surveillance was actually an offensive effort intended to undermine – or, more likely, give Washington the option of undermining – Beijing's or Moscow's ability to launch nuclear weapons, thus potentially generating crisis instability. By contrast, because Washington is apparently more confident in the survivability of its nuclear deterrent, cyber espionage directed against U.S. nuclear forces or C3I systems would be less likely to have the same result. Nonetheless, such operations would likely be of real concern to Washington and could, for example, be misinterpreted as a prelude to nuclear use by China or Russia.

Even if the two states involved in a crisis or conflict did not engage in any kind of deliberate cyber interference with one another's nuclear forces or C3I systems, one of them might wrongly conclude that the other had. Such a misperception, which could be the result of collateral effects or third-party action, could also induce escalation through crisis instability or misinterpreted warning.

A state that eschewed cyber operations of any kind against an opponent's nuclear forces or C3I systems might still launch such operations against adversary military networks involved exclusively in non-nuclear operations. If, because of design flaws, imperfect intelligence, or mistakes in execution, the malware used in such attacks spread and infected networks that were involved in nuclear

operations, the target might conclude that its nuclear forces or C3I systems were under deliberate cyberattack or cyber surveillance.

There could be collateral effects even if a state's networks for nuclear operations were entirely isolated; air-gapping (physically isolating one particular network from others) is, after all, not a cyber security panacea.[30] Moreover, achieving perfect isolation could prove difficult in practice.[31] To give but one reason, every nuclear-armed state, apart from the United Kingdom, has dual-use delivery systems, which can be used to deliver nuclear or non-nuclear weapons. Such delivery systems represent a potential point of contact between the C3I systems supporting nuclear operations and those supporting non-nuclear operations.

In practice, some nuclear-armed states – perhaps many or even all of them – have not tried to isolate their nuclear C3I systems. The United States, for example, has a number of dual-use C3I assets for communications and early warning that support both nuclear and non-nuclear operations.[32] Other nuclear-armed states, including China and Russia, may as well, but are less transparent.[33] Because the networks supporting dual-use C3I assets are likely to be connected directly to others involved in non-nuclear operations, there may be a particularly high risk of their being subject to collateral effects.

Catalytic warfare is a long-standing theoretical concern about a multipolar nuclear world that cyber capabilities could make all too real. During the Cold War, American strategists occasionally opined that China might try to take advantage of a U.S.-Soviet confrontation by firing nuclear weapons (most likely from submarines) at one or both of the superpowers in the hope that they would misattribute the origin of the attack and proceed to launch a nuclear war that would "weaken or destroy" each other.[34] Such fears were clearly absurd then. Armageddon was not in China's interests, even if it were only a bystander. While that remains even truer today, the advent of cyber warfare makes catalysis plausible, albeit as a result of inadvertence rather than deliberate action.

In peacetime, multiple nuclear-armed states may simultaneously prepare for conflict against the same adversary. Currently, for example, China, Russia, and North Korea all have incentives to try and penetrate the United States' nuclear forces and C3I systems. If a state with multiple adversaries detected malware in the networks supporting its nuclear forces, the identity of the perpetrator might not be immediately clear.[35] (The same would be true, of course, for attacks against other networks, but the consequences would be less significant.)

Especially in a conflict or crisis, the difficulty of resolving this uncertainty could have serious consequences. One key factor that affects the "quality of attribution" for cyber operations is time: as more time is spent on attribution, conclusions are likely to become more accurate and more confident.[36] As a corollary, "when high-level decisions . . . have to be made under pressure, the speed of

political developments may outpace the speed of the attribution process."[37] A crisis or conflict is one such circumstance. If a state found malware in its nuclear forces or C3I systems, then it might feel that it had no choice but to act on the assumption that its attacker was the other party involved in the contingency.[38] In February 1998, for example, the United States discovered a successful hack of military networks while preparing to bomb Iraq and, to quote White House official Richard Clarke, "assumed" that Baghdad was the culprit when, in fact, teenagers from Canada, Israel, and the United States turned out to be responsible.[39]

If a third party – and not the state's immediate adversary – were, in fact, to blame for a cyber intrusion, then catalytic escalation with the immediate adversary could result. The severity of the escalation pressures in this case is up for debate. On the one hand, awareness of any uncertainty associated with attribution might limit the forcefulness of any response. On the other, in contrast to the Cold War, when a catalytic strike by China would necessarily have been limited, a cyber intrusion might appear to be the precursor to an all-out damage-limitation attack, exacerbating the escalation risks.

A final difference between the cyber and noncyber weapons that can threaten nuclear forces and C3I systems is the much greater difficulty of limiting or otherwise cooperatively managing cyber capabilities. Strategic nuclear forces have long been subject to arms control, at least between the United States and the Soviet Union or Russia. Other relevant noncyber capabilities, including high-precision conventional munitions and antisatellite weapons, have generally not been subject to any form of international governance, and the technical and political challenges to managing them cooperatively are very real. These challenges, however, pale in comparison to those associated with governing cyber capabilities. Nonetheless, two ways forward present themselves.

First, states can and should act unilaterally to mitigate the risks. States should, for example, enhance their ability to prevent, detect, and mitigate the consequences of cyber interference with nuclear weapons and C3I systems and their associated supply chains. While much of the required effort here would be highly technical – finding vulnerabilities, scanning networks, and so forth – states should also consider whether they should change the way that their nuclear forces are postured and operated in order to help mitigate the consequences of what will inevitably be some degree of cyber vulnerability. To give but one example, any military that currently tracks the locations of its own mobile nuclear forces after dispersal could consider whether, to reduce the consequences of cyber espionage, it should stop doing so. Indeed, when a U.S. ballistic missile submarine is deployed on a deterrence patrol, its location is unknown except to submariners serving on that vessel. While this security precaution was developed long before the emergence of cyber warfare, it could help reduce the likelihood that cyber surveillance

of U.S. C3I networks might compromise the most survivable component of the United States' nuclear forces.

Restraint represents another form of unilateral risk reduction. In particular, states should adopt a consciously risk-averse approach to authorizing potentially escalatory cyber operations, particularly those that are targeted directly against nuclear forces or C3I systems, including dual-use networks. All of the escalation pathways outlined above, with the exception of false positives, involve a cyber operation by one state against another (even if the initiator could end up being a bystander to the subsequent escalation sequence). States, therefore, should put in place rigorous internal processes – if they do not already exist – to ensure that, in deciding whether to proceed with a potentially escalatory cyber operation, the strategic risks are fully considered and weighed against the potential intelligence and military benefits.

Conducting such assessments fairly and rigorously would likely prove difficult. One challenge would be deciding which cyber operations were "potentially escalatory" and so subjected to greater scrutiny. A second would be ensuring that low-probability but high-consequence escalation risks were not unduly discounted in comparison to more obvious and immediate military and intelligence benefits. Part of the solution should be to ensure that the assessment of escalation risks is not narrowly confined to the military or intelligence personnel responsible for proposing, planning, and conducting cyber operations. Such personnel are generally not trained in estimating – if an adversary detected a cyber operation – how threatening it might perceive the operation to be and how it might react. Rather, a broader cast of experts, including intelligence analysts who specialize in understanding foreign decision-makers, should be involved. In this context, this essay and other academic works hopefully have a role to play by identifying and raising awareness of the potential risks.

Ultimately, the authority to approve or reject a proposed cyber operation should rest with the senior officials who would be responsible for managing the real-world consequences of escalation. In the United States, for example, it should generally fall, if it does not already, to Senate-confirmed civilians. In the case of cyber interference that would directly affect the nuclear forces or C3I systems of another nation, however, the president should be the decision-maker. Again, this proposal is easier to suggest than to implement: for it to be effective, real buy-in from the bureaucracy would be required. Advisers would have to bring the decision-maker rapidly up to speed on complex technical details about the proposed operation and on the adversary's strategic culture and threat perceptions. Moreover, planners should develop two or more options that posed varying escalation risks – at least one of which did not involve any interference with nuclear forces or C3I systems – so that the decision-maker could properly assess any trade-offs between escalation risks and military and intelligence benefits.

Behavioral norms represent a more challenging but complementary pathway to reducing escalation risks. For example, states could agree, on a bilateral or multilateral basis, not to launch cyber operations of any kind against each other's nuclear forces or C3I systems. While such an agreement would not be verifiable in the traditional sense, it might nonetheless be enforceable: any state that considered launching a cyber operation in violation of the agreement would have to reckon with the possibility that the target (which would presumably be scanning its networks continuously) would detect the intrusion and respond in kind. In this way, deterrence could motivate compliance. To be sure, the challenges to reaching such an agreement would be daunting. In particular, it would likely be difficult to define what systems would and would not be covered by any prohibition, not least because of the existence of dual-use C3I assets. In the short term, however, more modest steps are possible. For example, states should reassure one another that any decision to launch a cyber operation against another state's nuclear forces or C3I systems, including dual-use networks, would be taken at the head of state or head of government level.

Norms are far from an ideal way to try to manage existential risks, but there is evidence that they can change behavior, including in cyberspace. In 2015, for example, President Barack Obama and President Xi Jinping agreed that neither of their states would engage in "cyber-enabled theft of intellectual property . . . with the intent of providing competitive advantages to companies or commercial sectors."[40] In 2018, the U.S. National Counterintelligence and Security Center assessed that Chinese cyber activity was taking place at "lower volumes" than before the agreement, and that it was mostly directed against "cleared defense contractors or IT and communications firms."[41] This statement indicates that China largely ceased conducting cyber activities for commercial gain, even if its compliance was not perfect. On balance, this experience suggests that trying to negotiate behavioral norms can be worth the effort, even if success is not guaranteed. Indeed, in the case of an agreement designed to prevent nuclear war, the incentives for compliance would be particularly strong.

If these suggestions seem to fall far short of the challenge presented by the potential risk of cyber interference with nuclear forces or C3I systems, it is because they almost certainly do. There is a profound mismatch between the importance of governing cyber capabilities and governments' (in)ability to do so. That said, modest steps may prove to have extrinsic value. For much of the Cold War, the idea that the United States and the Soviet Union might conduct inspections of one another's nuclear forces seemed far-fetched. But such inspections, which today involve counting the reentry vehicles emplaced on intercontinental ballistic missiles, were the culmination of a stop-start confidence-building process that began, after the Cuban missile crisis, with the modest first step of creating a hotline between the two superpowers. Political change in the Soviet Union

was unquestionably a necessary enabling condition for the breakthroughs of the late 1980s and early 1990s, but it might not have been possible to capitalize on such change had there not been an ongoing arms control process on which to build. There is no guarantee that an analogous process for managing cyber capabilities is possible. But if it is, it will inevitably begin with a modest first step.

ABOUT THE AUTHOR

James M. Acton holds the Jessica T. Mathews Chair and is Co-Director of the Nuclear Policy Program at the Carnegie Endowment for International Peace. His recent work includes the 2018 *International Security* article, "Escalation through Entanglement: How the Vulnerability of Command-and-Control Systems Raises the Risks of an Inadvertent Nuclear War."

ENDNOTES

[1] Richard Rhodes, *The Making of the Atomic Bomb* (New York: Simon & Schuster, 1986), 687.

[2] U.S. Department of Defense, *Nuclear Posture Review* (Washington, D.C.: Office of the Secretary of Defense, 2018), 56–57, https://media.defense.gov/2018/Feb/02/2001872886/-1/-1/1/2018-NUCLEAR-POSTURE-REVIEW-FINAL-REPORT.PDF.

[3] Defense Science Board, U.S. Department of Defense, *Task Force Report: Resilient Military Systems and the Advanced Cyber Threat* (Washington, D.C.: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, 2013), 6, https://nsarchive2.gwu.edu/NSAEBB/NSAEBB424/docs/Cyber-081.pdf. For theory and evidence supporting this conclusion, see Andrew Futter, *Hacking the Bomb: Cyber Threats and Nuclear Weapons* (Washington, D.C.: Georgetown University Press, 2018), chap. 2 and 4; Page O. Stoutland and Samantha Pitts-Kiefer, *Nuclear Weapons in the New Cyber Age: Report of the Cyber-Nuclear Weapons Study Group* (Washington, D.C.: Nuclear Threat Initiative, 2018), https://media.nti.org/documents/Cyber_report_finalsmall.pdf; Erik Gartzke and Jon R. Lindsay, "Thermonuclear Cyber War," *Journal of Cyber Security* 3 (1) (2017): 38–40; and Patricia Lewis and Beyza Unal, "Cyber Threats and Nuclear Weapons Systems," in *Understanding Nuclear Weapon Risks*, ed. John Borrie, Tim Caughley, and Wilfred Wan (Geneva: United Nations Institute for Disarmament Research, 2017), http://www.unidir.org/files/publications/pdfs/understanding-nuclear-weapon-risks-en-676.pdf.

[4] For the classic Cold War discussion of crisis instability, see Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, Mass.: Harvard University Press, 1960), chap. 9. For more recent treatments that stress limited nuclear use, see Caitlin Talmadge, "Would China Go Nuclear? Assessing the Risk of Chinese Nuclear Escalation in a Conventional War with the United States," *International Security* 41 (4) (2017): 50–92; and Michael S. Gerson, "No First Use: The Next Step for U.S. Nuclear Policy," *International Security* 35 (2) (2010): 35–39.

[5] See, for example, Futter, *Hacking the Bomb*, 117–125.

[6] James M. Acton "Escalation through Entanglement: How the Vulnerability of Command-and-Control Systems Raises the Risks of an Inadvertent Nuclear War," *International Security* 43 (1) (2018): 67–73. The focus of that work is dual-use C3I systems. However, because of their unique characteristics, cyber operations could also create misinterpreted warning, even in the case of C3I systems reserved exclusively for nuclear operations.

[7] U.S. Department of Defense, *Nuclear Posture Review*, 21.

[8] Deliberate interference could also create escalation pathways unique to cyberspace. For one such example, the "cyber commitment problem," see Gartzke and Lindsay, "Thermonuclear Cyber War," 41–45.

[9] For example, David C. Gompert and Martin Libicki, "Cyber Warfare and Sino-American Crisis Instability," *Survival* 56 (4) (2014): 7–22.

[10] Revisit times are likely to be longest for the most capable satellites, which can generally only be procured in small numbers because of their cost.

[11] Jon R. Lindsay, "Stuxnet and the Limits of Cyber Warfare," *Security Studies* 22 (3) (2013): 384.

[12] For example, David A. Fulghum, Robert Wall, and Amy Butler, "Cyber-Combat's First Shot: Israel Shows Electronic Prowess: Attack on Syria Shows Israel Is Master of the High-Tech Battle," *Aviation Week & Space Technology* 167 (21) (2007): 28–31.

[13] Steven M. Bellovin, Susan Landau, and Herbert Lin, "Limiting the Undesired Impact of Cyber Weapons: Technical Requirements and Policy Implications," in *Bytes, Bombs, and Spies: The Strategic Dimensions of Offensive Cyber Operations*, ed. Herbert Lin and Amy Zegart (Washington, D.C.: Brookings Institution Press, 2018), 274–276.

[14] Other causes of collateral damage, such as planners' misidentifying a civilian facility as a military one, are less easy to quantify.

[15] Unmanned aerial vehicles capable of both intelligence collection and offensive operations are an exception because they can be used for surveillance ahead of a decision to employ them for offensive purposes. It seems unlikely, however, that one nuclear-armed state would use such platforms against another in this way in peacetime.

[16] Lin and Zegart, "Introduction," in *Bytes, Bombs, and Spies*, 6.

[17] U.S. Department of Defense, *Nuclear Posture Review*, 23.

[18] U.S. thinking about the requirements for force survivability can be explained only by the belief that Russia might conduct counterforce strikes.

[19] Christopher Clary and Vipin Narang, "India's Counterforce Temptations: Strategic Dilemmas, Doctrine, and Capabilities," *International Security* 43 (3) (2018/2019): 7–52.

[20] See, for example, Keir A. Lieber and Daryl G. Press, "The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence," *International Security* 41 (4) (2017): 9–49; Charles L. Glaser and Steve Fetter, "Should the United States Reject MAD? Damage Limitation and U.S. Nuclear Strategy Toward China," *International Security* 41 (1) (2016): 63–70; and Austin Long and Brendan Rittenhouse Green, "Stalking the Secure Second Strike: Intelligence, Counterforce, and Nuclear Strategy," *Journal of Strategic Studies* 38 (1–2) (2015): 38–73.

[21] For discussions of Chinese and Russian concerns, see, for example, Alexei Arbatov, Vladimir Dvorkin, and Sergey Oznobishchev, *Non-Nuclear Factors of Nuclear Disarmament: Ballistic Missile Defense, High-Precision Conventional Weapons, Space Arms* (Moscow: IMEMO RAN, 2010), https://www.files.ethz.ch/isn/144178/10002.pdf; and Fiona S. Cunningham and M. Taylor Fravel, "Assuring Assured Retaliation: China's Nuclear Posture and U.S.-China Strategic Stability," *International Security* 40 (2) (2015): 15–23.

[22] Paul Bracken, "The Cyber Threat to Nuclear Stability," *Orbis* 60 (2) (2016): 197–200.

[23] Tong Zhao and Li Bin, "The Underappreciated Risks of Entanglement: A Chinese Perspective," in *Entanglement: Russian and Chinese Perspectives on Non-Nuclear Weapons and Nuclear Risks*, ed. James M. Acton (Washington, D.C.: Carnegie Endowment for International Peace, 2017), 62.

[24] Alexey Arbatov, Vladimir Dvorkin, and Petr Topychkanov, "Entanglement as a New Security Threat: A Russian Perspective," in *Entanglement*, ed. Acton, 13. They go on to add, however, that "radio channels for communicating with and controlling satellites–especially missile early-warning assets–are more vulnerable."

[25] Sergey Rogov, "Disarmament and Deterrence–Bridging the Divide," remarks at the 5th EU Non-Proliferation and Disarmament Conference, Brussels, Belgium, November 3, 2016, https://www.iiss.org/-/media/images/dialogues/eunp/eunp-2016/documents/disarmament-and-deterrence---bridging-the-divide_-sergey-rogov-_-iiss.ashx.

[26] "Cyber Security Units to Protect Russia's Nuclear Weapons Stockpiles," RT, October 17, 2014, https://www.rt.com/news/196720-russia-missile-forces-cybersecurity/.

[27] For example, David Hoffman, "'I Had a Funny Feeling in My Gut,'" *The Washington Post*, February 10, 1999, http://www.washingtonpost.com/wp-srv/inatl/longterm/coldwar/shatter021099b.htm.

[28] Martin C. Libicki, *Brandishing Cyberattack Capabilities* (Santa Monica, Calif.: RAND Corporation, 2013), 22–23, https://www.rand.org/content/dam/rand/pubs/research_reports/RR100/RR175/RAND_RR175.sum.pdf.

[29] For a dated example, see Daniel Ellsberg, *The Doomsday Machine: Confessions of a Nuclear War Planner* (New York: Bloomsbury, 2017), 53–57. It bears emphasizing that technological developments have massively reduced the likelihood of a communications blackout and that the United States no longer predelegates launch authority. Rather, what's significant in Ellsberg's account is the predilection of the military officers he interviewed to regard a communication blackout as an attack.

[30] Martin C. Libicki, *Cyberspace in Peace and War* (Annapolis, Md.: Naval Institute Press, 2016), 49–50.

[31] Ibid.

[32] Acton "Escalation through Entanglement," 63–65.

[33] Ibid., 65–66 and 78–80.

[34] Richard Rosecrance, *Strategic Deterrence Reconsidered*, Adelphi Paper 116 (London: The International Institute for Strategic Studies, 1975), 33.

[35] The literature on catalytic escalation in cyberspace emphasizes the somewhat different case of a third party that wants to spark a conflict between two others. See, for example, Futter, *Hacking the Bomb*, 118; and Libicki, *Cyberspace in Peace and War*, 315.

[36] Thomas Rid and Ben Buchanan, "Attributing Cyber Attacks," *Journal of Strategic Studies* 38 (1–2) (2015): 32.

[37] Ibid.

[38] Indeed, in wartime, states have to make such assumptions since it is simply not possible to attribute every incoming attack.

[39] William M. Arkin, "Sunrise, Sunset," *The Washington Post*, March 29, 1999, http://www.washingtonpost.com/wp-srv/national/dotmil/arkin032999.htm. Relatedly, there is also evidence of states' attempting to collect intelligence about adversaries' capabilities by targeting third parties. Ben Buchanan, *The Cybersecurity Dilemma: Hacking, Trust, and Fear Between Nations* (Oxford: Oxford University Press, 2016), 69–71.

[40] Office of the Press Secretary, The White House, "Fact Sheet: President Xi Jinping's State Visit to the United States," September 25, 2015, https://obamawhitehouse.archives.gov/the-press-office/2015/09/25/fact-sheet-president-xi-jinpings-state-visit-united-states.

[41] U.S. National Counterintelligence and Security Center, *Foreign Economic Espionage in Cyberspace, 2018* (Washington, D.C.: Office of the Director of National Intelligence, 2018), 7, https://www.dni.gov/files/NCSC/documents/news/20180724-economic-espionage-pub.pdf.